The Use of a Dimeric Four-helix Bundle as a Model to Study Protein Folding and Design

A senior honors thesis

Presented in partial fulfillment of the requirements for graduation *with distinction* in Biochemistry in the undergraduate colleges of The Ohio State University

by

Sanjay Baju Hari

The Ohio State University
June 2008

Project Advisor: Thomas J. Magliery, Departments of Chemistry and Biochemistry

Committee:
Dr. Thomas J. Magliery
Dr. Venkat Gopalan
Dr. Ross E. Dalbey

Acknowledgements

Thomas J. Magliery, project advisor, Ohio State University departments of Chemistry and Biochemistry
Venkat Gopalan, oral exam committee, Ohio State University department of Biochemistry
Ross E. Dalbey, oral exam committee, Ohio State University department of Chemistry

Jason J. Lavinder, OSBP

Chang Byeon, Ohio State department of Chemistry

Jinjin Zhang, OSBP, laboratory of Charles E. Bell

Jeremy Beck, department of Chemistry, laboratory of Christopher Hadad

Funding

Ohio State Department of Biochemistry Summer Undergraduate Research Program (SURP) fellowship

Ohio State College of Arts and Sciences research scholarships

College of Biological Sciences Mayers Summer research fellowship

College of Biological Sciences Dean's Undergraduate research scholarships

Abstract

Fifty years after the work of Anfinsen, we are still unable to definitively determine the tertiary structure of proteins from their sequences. While it is true that recent computational studies have shown promising results, it is still very difficult to conclude whether a given sequence will fold at all, let alone predict its conformation. Experimental approaches to the protein folding problem allow us to improve upon computational models by obtaining physical data on how specific residue mutations affect structure, function, and stability.

The four-helix bundle rop was chosen for our studies. Rop is a small protein with an abundance of high-resolution data, making it an ideal choice for site-directed mutagenesis experiments. Additionally, an *in vivo* screen which uses GFP as a reporter exists for rop.

We first set out to create a cysteine-free variant of rop that behaves like the wild-type in structure, function, and stability. After testing several mutants that replace the two cysteines with similar residues, it was found that the A38/V52 variant was functionally active, expressed well in solution, and possessed a high degree of α-helical content. ^{15}N-HSQC NMR data of this A38/V52 mutant showed a peak dispersion that was very similar to wild-type. Preliminary x-ray crystallographic data indicates that this cys-free mutant deviates from wild-type by 0.37 Å.

A new method for purifying and screening protein library variants on a high-throughput scale was developed. The purification method takes advantage of recent developments in high-throughput protein manipulation tools, and the screening method is able to inexpensively screen almost one hundred variants at once for stability.

Two single-chain rop constructs were designed with the intent to create a single-chain monomer that mimicked the structure and function of wild-type rop, which is normally a dimer. Both constructs were found to be functionally active. Further characterization is in progress.

THIS PAGE INTENTIONALLY LEFT BLANK

Introduction

Feats of folding

Christian Anfinsen's groundbreaking work on ribonuclease in the 1950's definitively linked protein sequence to tertiary structure.[1] Denaturing the protein fully with urea ensured that no other factors would influence its subsequent refolding. Upon removal of the denaturant and addition of a catalytic amount of reducing agent, the RNase was able to assume its original conformation. This established that the amino acid sequence of a protein is the sole determinant of its final conformation in solution. Hence, each residue presumably contributes in some way to a protein's biophysical properties, including structure and stability. Anfinsen's work certainly advanced our understanding of protein science; knowing that nature only needs a residue sequence to generate a tertiary structure meant that the goal of predicting structures was a tangible one. However, it also meant that a host of new questions would have to be answered.

For example, how does residue sequence govern protein folding, and how does this folding occur in a finite period of time? Proteins can theoretically assume an almost infinite number of conformations, yet they could not possibly try all of them. This dilemma is commonly known as Levinthal's paradox, named after Cryus Levinthal who posited the issue in the 1960's.[2] In fact, proteins in nature fold in fractions of a second, likely through a folding "funnel" which iteratively leads to structures in their lowest energy conformations.[3] This begets another question of how much information is encoded in any particular residue. Do some positions in a sequence "contribute" more to a structure than others? Perhaps a more immediate question is, what are the forces that act upon amino acids during folding? These questions are encompassed in what is commonly known as the protein folding problem. Active research in these areas is done with the expectation to one day be able to accurately predict protein structures

from their respective sequences. Although considerable progress has been made towards answering these questions, much work still remains.

The energetics of protein folding are complex and beyond the scope of this text. However, a brief overview is reasonable to explain other topics covered. The hydrophobic core of a globular protein is central to its folding.[4] Nonpolar sidechains of hydrophobic residues seek to remove themselves from water to obtain a favorable, spontaneous change in free energy by interacting with other nonpolar groups. A group of these nonpolar sidechains which bury themselves within the protein constitutes the hydrophobic core. The resultant van der Waals forces inside this hydrophobic core appear to contribute to the protein's stability.[5, 6] Surprisingly, a protein is only marginally stable; the large, opposing factors of enthalpy and entropy virtually cancel out each other in a well-folded protein to yield a very small free energy value.[4]

Approaches

There are two principle methods of studying protein folding today: experimental and computational. The latter has emerged as a powerful tool in recent years.[7] Dahiyat and Mayo[8] reported the computational design of a protein sequence not found in nature whose NMR structure closely resembled its predicted structure. This was accomplished by using algorithms to find a sequence that would fold into a desired backbone configuration, in this case a ββα motif. Harbury et al. used computational methods to design right-handed alpha-helical structures not found in nature.[9] Instead of using a rigid backbone, though, a different approach was used to take into account both main chain flexibility and side chain packing interactions. More recently, Kuhlman et al. used the program Rosetta to computationally design a novel fold using energy minimization and optimization.[10] The resultant protein crystal structure deviated from the

predicted structure by 1.2 Å.

Experimental methods in protein science are based largely on analyzing the structural effects of residue mutations, or site-directed mutagenesis. As mentioned before, the folding of proteins is a delicate energetic balancing act. It is still difficult for even the best computational algorithms to calculate large enthalpic and entropic values to such a high degree of accuracy that the resulting free energy value does not have a larger margin of error than the value itself. Hence, studying proteins directly is an excellent way to investigate the driving forces of protein folding, which could help refine computational methods. Hecht and Sauer's report over two decades ago illustrated the different consequences of mutating solvent exposed residues versus those within the hydrophobic core.[11] The authors made over fifty point mutations of λ repressor and found that, while surface residue mutations did not substantially alter its overall folding and stability, they did seem to adversely affect its ability to bind DNA. Conversely, mutating core residues did appear to affect packing and folding. Five years later, Lim and Sauer reported the successful repacking of λ-repressor's hydrophobic core.[12] Their experiments revealed that nonpolar residues could largely only be replaced by like residues (Ala, Val, Ile, Leu, Met, Phe). Cysteine, threonine, and serine were also mildly tolerated in some locations, which was thought to be due to their hydrogen bonding abilities. Later, the authors characterized several λ repressor variants with hydrophobic core mutations.[13] These mutations were intentionally "disruptive," conferring a drastic change in either core volume or hydrophobic character. They found that these mutations often induced changes in both structure and stability.

Libraries

Protein folding studies have been aided tremendously by advances in library synthesis and screening. Libraries are already heavily used in small molecule and peptide drug lead

discovery[14] as well as for finding novel protein binding partners.[15] The latter aim has been facilitated greatly by solid phase peptide synthesis,[16] which has allowed the automated synthesis of very large libraries. In the realm of molecular biology, phage display has proven to be an excellent method to be able to quickly screen peptide libraries and establish a "physical linkage between phenotype and genotype,"[17] which greatly expedites sequence identification. Libraries of larger proteins can be cloned directly into vectors and transformed into cells for screening.[18] This shift from direct synthesis to in-cell synthesis via genetic transcription and translation was made possible by the use of nucleotide mixes in oligonucleotide synthesis.[19, 20] By introducing an equimolar mix of more than one nucleotide type into a reaction vessel during a synthesis step, the resultant oligonucleotides are extended by one of up to four different bases. Hence, the genes made using this oligonucleotide may differ in this region. Using nucleotide mixes in several locations rapidly increases the diversity of the subsequent library. Doing so in a rational fashion allows the incorporation of only certain types of residues (hydrophobes, for example). This has obvious applications in experiments which require such stringency: core repacking, for example.[21] Libraries that randomize all or part of the hydrophobic core can yield extremely useful insight into the underpinnings of protein folding.

Screening

The abundance of library construction methods has facilitated the rapid synthesis of diverse libraries.[22] The bottleneck now lies in screening these libraries on a high-throughput scale; traditional methods such as circular dichroism (CD) and NMR spectroscopies, while well-studied, are simply unable to process a large number of proteins in a reasonable time frame. Both require preparations on the order of 1 L cell culture, and some versions of NMR spectroscopy entail atomic labelling. Characterizing ten variants by these methods is difficult, let alone libraries that are

several orders of magnitude larger. A number of methods have emerged to meet this need for high-throughput characterization. Crystallography has seen steady progress, permitting automated, high-density microtray preparation and micron-size crystal diffraction.[23] Additionally, crystallographic software has improved dramatically; some programs now permit automatic molecular replacement and molecule building.[24] This can be especially useful for characterizing variants whose wild-type crystal structures have already been solved.

Edgell and colleagues[25] established fluorescence as a viable option for high-throughput characterization by using a titrating fluorometer to monitor the unfolding of eglin c mutants. This dual-channel fluorometer slowly combined chemical denaturant with protein while measuring intrinsic fluorescence. The throughput was reported to be approximately 20 min per protein. A major drawback to this method was its reliance on side chain aromatic groups to measure fluorescence. Thus, proteins without Tyr or Trp residues could not be screened.

This work was followed by the development of ThermoFluor®, a method used to assess ligand binding by thermal denaturation of protein variants in the presence of a fluorescent dye.[26] Upon exposure to the protein's hydrophobic core, the dye experiences an increase in quantum yield.[27, 28] Hence, the unfolding of the protein can be monitored by fluorescence. Unlike the method described by Edgell, this was protein-independent because of its use of a reporter dye instead of intrisic fluorescence. Particularly noteworthy of ThermoFluor® was that it used a CCD camera to analyze separately each well of a 96-well plate, making it possible to screen approximately 3000 compounds in a timely fashion.

Recently, a Thermofluor®-based method has emerged to serve a host of functions. Lo et al. used an RT-PCR machine and a commercially available dye, SYPRO Orange, to assess the ligand binding of BACE1 protein.[29] The methodology behind this technique is identical to that of

Thermofluor: the thermal unfolding of a protein is reported by a hydrophobic dye. Ericsson et al. used this new method to find buffers and additives that would stabilize proteins in order to increase their propensity to crystallize.[30] The same was true for Yeh et al. with membrane proteins.[31] Vedadi et al. aimed for the same result, but used different ligands instead of conditions.[32]

De novo design

An important tool in advancing our understanding of protein folding is *de novo* design. It allows us to benchmark how well we can correlate sequence to tertiary structure at our present level of understanding. Regan and DeGrado used a bottom-up approach to design a non-natural protein with a four-helix bundle motif.[33] This specific motif is popular in *de novo* studies in part because its α-helical content can be measured directly by CD spectroscopy. In an iterative approach taken in that report, the authors first designed peptides that folded into α-helices. Next, they designed a loop to connect two of these helices. In the final phase, the dimers were connected by another identical loop to yield a self-assembling four-helix bundle.

Later, the development and characterization of the non-natural protein α_2D marked the third generation of designed, dimerizing four-helix bundles from the DeGrado group.[34] Initial characterization of this protein indicated a well-folded structure that was high in α-helical content. Additionally, α_2D denatured cooperatively upon heating. From this information, one would expect the protein to have self-associated in a nativelike manner. However, when the solution structure was determined,[35] it was discovered that α_2D in fact dimerized as a "bisecting U," a conformation not ordinarily found in nature.

Rop as a model protein

Cesareni et al. reported that a 63-residue protein was able to inhibit the binding of a specific primer RNA precursor to the ColE1 origin, thereby controlling the level of ColE1 replication in *E.*

coli.[36] They dubbed this protein Repressor of Primer, or rop. A communication by Banner et al. noted that rop was likely a dimer, and that the protein appeared to crystallize under a variety of conditions.[37] The high-resolution structure was reported shortly thereafter;[38] rop was an antiparallel, dimeric four-helix bundle composed almost entirely of α-helical content. The solution structure reported later was in agreement with the crystal structure.[39]

The large amount of high-resolution data for rop makes this protein an ideal choice for studying four-helix bundles. The first mutagenesis experiments performed on rop were done to find its RNA-binding site.[40] During their studies, the authors found that rop is surprisingly resiliant to surface mutations, despite being a relatively small protein.

"Repacking" a core generally entails replacing the residues inside the hydrophobic region of a well-folded protein. This is done in an effort to maintain the structural and stability characteristics of wild-type, but using a simplified inner region. Repacking studies on rop were performed first by Munson et al.[41] The authors treated rop's core as eight layers of residues stacked by the protein's α-helical structure. Constructs were made in which part or all of the core was redesigned such that each helix contributed only either Leu or Ala to the core. These new variants were found to show similar results to wild-type rop upon biophysical characterization.

Converting the normally dimeric rop into a single-chain monomer has also been an issue of interest. As Predki et al. reported, a monomeric form would simplify mutational studies because variations on one side of the protein would not necessarily be conferred to the other side.[42] Also, the successful design of a native-like, single-chain variant would yield the ability to rigorously compare the dimeric and monomeric interfaces. Currently, one can examine separately monomers and dimers to see how they differ, but any deductions made would be inherently flawed because they compare different proteins. However, by using a single-chain version of essentially the same

protein, one would be able to more accurately assess how residue mutations affect a dimeric interface versus a monomeric one.

In the first attempt at single-chain rop, glycine chains were used to connect the helices.[42] The authors found that loop length was very important in preventing oligomerization, and that constructs of adequate loop lengths displayed cooperative unfolding. Another group[43] continued the study by choosing and optimizing appropriate loops from the Protein Data Bank. From these experiments, they concluded that loop composition is also extremely important in creating a well-folded structure.

To replace the difficult gel-shift assays used to determine if a rop variant is functionally active, Magliery and Regan developed a cell-based assay that uses Green Fluorescent Protein as a reporter.[44] As noted before, the function of rop is to regulate the copy number of ColE1 plasmid. By cloning GFP into a plasmid that uses a ColE1 origin, the authors were able to put rop directly in control of the amount of GFP expressed. A functionally active variant would limit plasmid replication, thereby limiting fluorescence caused by GFP. On the other hand, a misfolded rop variant would not be able to control plasmid replication; a high amount of GFP would result. The application of this screen in high-throughput studies is apparent; one could feasibly assay the function of hundreds of rop variants on a single semisolid agar plate.

Objectives

1: Synthesis and characterization of a cysteine-free rop

The cysteines present at positions 38 and 52 on the wild-type rop monomer present a challenge in library synthesis. Normally not a problem *in vivo* due to the reducing environment of *E. coli*, these cysteines tend to form irreversible disulfide bonds within the hydrophobic core across the dimer. This is especially problematic for destabilized library variants, which tend to

aggregate upon disulfide formation. Thus, a cysteine-free rop variant that closely resembles the wild-type in structure and function would be extremely useful as a template for subsequent libraries.

2: Development and application of a novel high-throughput calorimetric assay

Library screening performed *in vitro* entails two distinct challenges: isolating each protein variant to homogeneity and finding a method to assay these variants. In addition, both steps must be performed within a reasonable period of time. Unlike binding assays such as beads and phage display, assays that probe stability are particularly challenging because they require relatively large amounts of protein. Thus, one must be able to purify an adequate amount of protein from several orders of magnitude less culture than ordinarily used. Regarding the assay itself, it should ideally be rapid and simple to perform. Additionally, it would be inexpensive and require only readily accessible equipment.

3: Design and synthesis of single-chain rop variants

As noted before, the sixty-three-residue rop monomer forms an antiparallel homodimer to produce a well-packed hydrophobic core. By synthetically moving and adding loops to connect each of rop's four helices, it can be converted from a dimer to a single-chain monomer. While rop monomers have previously been made by several groups, they were not done with the specific intent to conserve the structure, stability, and function of wild-type. The variants aimed to be synthesized here would be designed with three unique qualities: they would attempt to mimic the charge distribution seen in wild-type rop, be functionally active, and contain no cysteines.

Methods

Objective 1: Synthesis and characterization of a cysteine-free rop

Gene construction (performed by C. Byeon). Oligonucleotides were purchased (Sigma-Genosys) to synthesize genes with point mutations at positions 38 and 52 of wild-type rop. These mutation combinations were Ala/Ala, Ser/Ser, Thr/Thr, and Val/Val. The genes were constructed using the general method described by Stemmer.[45] Briefly, the purchased oligonucleotides were designed to overlap in a complementary fashion by approximately 25 base pairs. Upon PCR annealing and extension, these oligonucleotides should join and extend to form longer strands of DNA. With enough overlapping fragments, entire genes are formed. The gene of interest is amplified with flanking primers that also contain desired restriction and protease sites. The oligonucleotides used to form these rop mutants are shown in Table 1. Primers 1 and 2 were used for all genes, while the third and fourth primers were specific to each variant. Bold lettering indicates the point mutation.

Table 1. *Primers used for cys-free rop mutant gene construction*

Primer	Sequence (5'-3')
1	ATGACCAAACAGGAAAAAACCGCCCTTAACA TGGCCCGCTTTATCAGAAGCCAGACATTAAC
2	CTGCCTGTTCATCCGCGTCCAGCTCGTTGAGT TTCTCCAGCAGCGTTAATGTCTGGCTTCTG
3 Ala	GACGCGGATGAACAGGCAGATATC**GCG**GAA TCGCTTCACGACCACGCTGATGAGCTTTACCG
3 Ser	GACGCGGATGAACAGGCAGATATT**AGC**GAA TCGCTTCACGACCACGCTGATGAGCTTTACCG

3 Val	GACGCGGATGAACAGGCAGATATT**GTG**GAAT CGCTTCACGACCACGCTGATGAGCTTTACCG
4 Ala	TCAGAGGTTTTCACCGTCATCACCGAAACGCG CGAG**CGC**ACTGCGGTAAAGCTCATCAGCG
4 Ser	TCAGAGGTTTTCACCGTCATCACCGAAACGCG CGAG**GCT**GCTGCGGTAAAGCTCATCAGCG
4 Val	TCAGAGGTTTTCACCGTCATCACCGAAACGCG CGAG**CAC**ACTGCGGTAAAGCTCATCAGCG
PMR pro	AATAATCCATGGCGCATCATCACCATCATCAC GGCGGTGAAAACCTGTATTTTCAGGGCACCAAACAGGAAAAAAC
PMR term	AATAATGGATCCTCAGAGGTTTTCACCGTC
PAC pro	AATAATAATCATATGACCAAACAGGAAAAAAC
PAC term	AATAATGGTACCTCAGAGGTTTTCACCGTC

Cloning (performed by C. Byeon). Each gene was amplified with two sets of primers to form "screening" genes that were used for the *in vivo* functional assay, and "expression" genes that contained N-terminal TEV- and His_6-sites. Screening genes (without TEV or His_6 site) and pAC vector were digested with NcoI and BamHI (NEB), and expression genes and pMR vector were digested with NdeI and BanI. Overnight ligation at 16 ºC was followed by transformation into electrocompetent DH10B, and recovered cells were plated and incubated overnight. Picked colonies were grown in rich media, and the extracted DNA (Qiagen Miniprep) was sequenced (OSU PMGF).

Screening (performed by C. Byeon). Genes in pAC vector were transformed into electrocompetent DH10B already containing the screening plasmid pUCBAD-GFPUV.[44] Cultures were plated onto semisolid LB media containing Amp/Kan/IPTG (0.1 mM)/arabinose (0.0005% w/v) and incubated overnight at 42 ºC. Fluorescent cells were visualized by UV light.

Protein purification. Rop variants in pMR vectors were overexpressed in BL21(DE3) from colonies or glycerol stocks in 1 L 2YT media. After induction to 0.1 mM IPTG at log phase, the cells were incubated for an additional 3 h at 37 ºC or overnight at 30 ºC. Frozen, harvested cell pellets were resuspended in 25 mL lysis buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 10 mM imidazole), mixed with 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 5 μL Dnase I, 5 μL Rnase A, 0.1% Triton X-100, and 0.3-1 mg/mL HEW lysozyme, and allowed to incubate on ice for 30 min. The solution was then sonicated at half power for 3x30 s with 2 min on ice in between pulses. The cleared lysate was mixed on ice with 1 mL Ni-NTA (Qiagen) for 1 h. The slurry was poured into a column, washed with wash buffer (lysis buffer with 20 mM imidazole), and the protein was eluted with elution buffer (lysis buffer with 250 mM imidazole). TEV protease[46] was added twice to the protein and incubated each time overnight at RT or 3 h at 30 ºC. The solution was desalted with PD-10 columns (GE Amersham) and mixed with 1 mL Ni-NTA. After incubating on ice for 1 h, the solution was added to a column, and the flow-through was collected to yield cleaved protein. This protein was concentrated through a YM-3 filter (Millipore) and exchanged into appropriate buffer for subsequent analysis.

CD spectroscopy. Data was collected on an Aviv 202 Circular Dichroism Spectrometer. Rop variants were analyzed at 50 μM monomer, as measured by UV absorption at 280 nm, and in CD buffer (50 mM sodium phosphate pH 6.3, 300 mM NaCl). Thermal denaturations were acquired at 1 °C min^{-1}, 25 to 90ºC, at 222 nm. Urea denaturations were performed in the same conditions

but using varying concentrations of urea, with spectra acquired after equilibrating 24-48 h at RT, at 222 nm.

NMR. Rop was purified as above, but using minimal media with ^{15}N-labeled NH_4Cl as the nitrogen source. The purified protein was exchanged into CD buffer, concentrated, and finally mixed with 10% D_2O directly before analysis. ^{1}H- and 15N-HSQC spectra were acquired on a 600 MHz Bruker NMR with inverse probe using standard Lewis E. Kay pulse sequences.

Crystal growth. Rop was purified as above, but was followed by a gel filtration purification using a Superdex 75 10/300 GL column (GE Tricorn). The protein was then exchanged into mild buffer (10 mM PIPES pH 6.3, 50 mM NaCl) and concentrated, if necessary, to approximately 2 mg/mL as determined by Bradford assay. Sitting drop and hanging drop trials were performed. Plates were obtained from both ammonium sulfate and MPD. Large prisms were obtained from hanging drop wells using 1-mL reservoirs containing 25-30% methanol, MES (100 mM pH 5.7-5.9), NaCl (300 mM), and glycerol (10%). Rhomboidal prisms appeared overnight at 22 ºC and grew for 3-5 days.

Crystal analysis. Diffraction data was collected on a Rigaku R-AXIS IV++ at –160 ºC and a detector distance of 200 mm. Images were collected at 0.5º increments with exposure times of 2 min. Integration was performed by d*TREK.[47] Molecular replacement modelling and building was done using the Phenix suite[24] and Phaser.[48] Refinement was performed using Coot[49] (including REFMAC5[50]) and Phenix.refine[24]

Objective 2: Development and application of a novel high-throughput calorimetric assay

Protein purification. 2YT rich media (1.5 mL) was inoculated with rop variants from colonies of DH10B(DE3) and grown overnight to saturation in 96-well deepwell plates (USA Scientific). Cultures were diluted to log-phase (OD_{600} ~0.8) and a volume of 2 mL, induced with 10 μM

IPTG, and grown overnight at 30 °C (described by Chae et al.)[51]. Harvested cell pellets were frozen at –80 °C for at least 1 h, then resuspended in lysis buffer. Lysis was performed by mixing lysates with DNase, RNase, lysozyme, and PopCulture reagent (Novagen), then incubating at RT for 30 min. Cleared lysates were mixed with 50 μL Ni-NTA magnetic agarose beads (Qiagen) and incubated at RT for 1 h. The beads were washed twice with wash buffer, then resuspended in 25 μL lysis buffer. On-column TEV cleavage was performed by adding 11 mM βME and 0.5 μL TEV protease and incubating at RT overnight with occasional agitation during the first few hours. The magnetic beads were removed, and the cleaved, purified protein was ready for further analysis.

Screening. Protein solution (19 μL) was mixed with 1 μL Sypro Orange (Invitrogen) to achieve a final concentration of 15x dye from a stock of 5000x. Melts were performed in RT-PCR plates on a Bio-Rad iCycler iQ Real-Time Detection System (OSU PMGF) set to increase temperature at a rate of 1 °C-min^{-1}. Excitation/emission wavelengths were set to 490/575 nm.

Objective 3: Design and synthesis of single-chain rop variants

Gene construction. Two single-chain mutants, MPP and PMP rop, were designed using the topologies shown in Figure 1.[43] Loops were designed by first principles in an effort to contain a reasonable amount of flexibility and solubility. PCR was performed as described in Objective 1. Oligonucleotide design for PMP rop used the method described by Gao.[52] Tables 2 and 3 show the oligonucleotides used for MPP and PMP rop syntheses, respectively. Flanking primers at the 5' end containing AflIII restriction sites for PMR and PAC vectors were kindly provided by J. Lavinder.

Figure 1. *Topologies of MPP (left) and PMP rop*

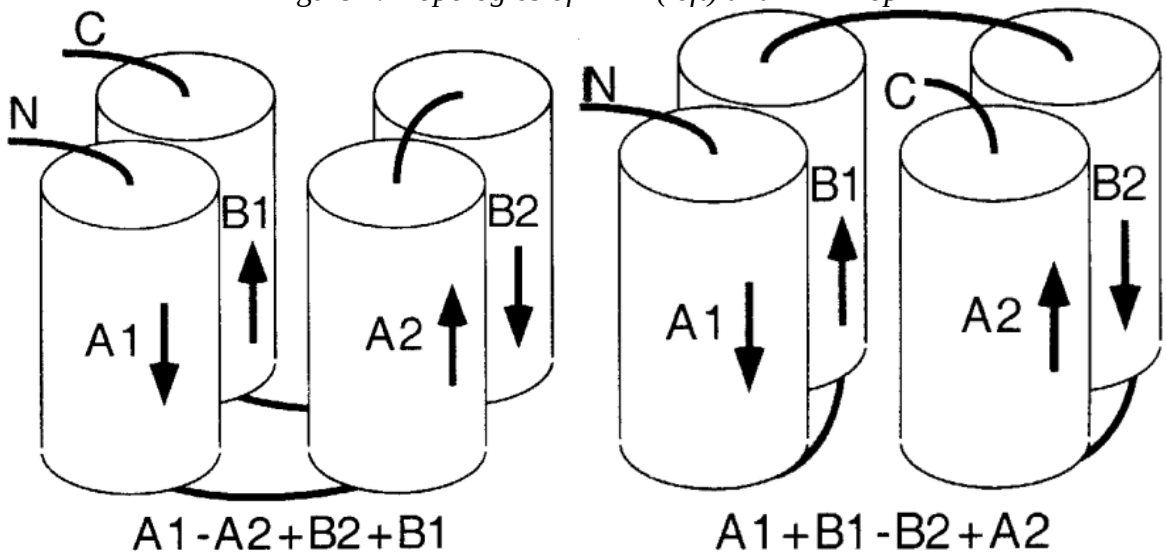


Table 2. *Primers used for single-chain MPP rop mutant gene construction*

Primer	Sequence (5'-3')
1	ATGACCAAACAGGAAAAAACTGCTCTCAACATGGCGCGTTTCAT
2	TCTCAACATGGCGCGTTTCATTCGTAGTCAGACCCTGACACTGCT
3	TAGTCAGACCCTGACACTGCTGGAGAAGCTGAACGAACTGGATGC
4	GAAGCTGAACGAACTGGATGCTGATGAACAGGCAGATATCGCGGA
5	GAACAGGCAGATATCGCGGAGAGTCTTCACGACCATGCGGACGAA
6	CTTCACGACCATGCGGACGAATTGTATCGAAGTGTGCTGGCGCGGTT
7	GAAGTGTGCTGGCGCGGTTCGATGGAGATGATGAGCAGGCCGACATA
8	GCGTGATCATGTAGACTTTCAGCTATGTCGGCCTGCTCATCATCT
9	AACACCGAACGGTACAGTTCATCTGCGTGATCATGTAGACTTTCAGC
10	TTATCTCCATCAAAGCGCGCCAACACCGAACGGTACAGTTCATCT
11	CATATTCAGCGCCGTTTTTTCCTGTTTATCTCCATCAAAGCGCGCC
12	TTGCGACCGTATAAACCGTGCCATATTCAGCGCCGTTTTTTCCTG
13	CAGTTTTTCAAGCAGAGTCAGTGTTTGCGACCGTATAAACCGTGC
14	TTATTTCAACTCATTCAGTTTTTCAAGCAGAGTCAGTGT
PMR-3'	AATAATGGATCCTTATTTCAACTCATTCAGTTTTTCAAG
PAC-3'	AATAATGGTACCTTATTTCAACTCATTCAGTTTTTCAAG

Table 3. *Primers used for single-chain PMP rop mutant gene construction*

Primer	Sequence (5'-3')
1	ATGACCAAACAGGAGAAAACGGCGCTTAACA
2	GCTCCGAATGAATCGTGCCATGTTAAGCGCCGTTTTTTCCTGTTT
3	GGCACGATTCATTCGGAGCCAGACCCTTACTCTCCTAGAAAAACT
4	TGCCGCTGTTCAGCTCGTTCAGTTTTTCTAGGAGAGTAAGGGTCT
5	AACGAGCTGAACAGCGGCACGAAACAGGAAAAGACGGCACTGAAC
6	TTGACTACGAATAAAACGCGCCATGTTCAGTGCCGTCTTTTCCTGT
7	GGCGCGTTTTATTCGTAGTCAAACCCTGACGCTGCTGGAGAAGCT
8	TGTTCATCCGCATCCAGTTCATTCAGCTTCTCCAGCAGCGTCAGG

9	TGAACTGGATGCGGATGAACAAGCAGACATCGCGGAAAGCCTGCA
10	GGTAAAGTTCGTCAGCGTGGTCATGCAGGCTTTCCGCGATGTCT
11	TGACCACGCTGACGAACTTTACCGGAGTGTGCTTGCACGTTTTGA
12	TGTCCGCCTGCTCGTCATCTCCATCAAAACGTGCAAGCACACTCC
13	GATGACGAGCAGGCGGACATAGCGGAGAGTTTGCATGATCACGCC
14	AGCACACTACGGTACAGCTCATCGGCGTGATCATGCAAACTCTCC
15	TGAGCTGTACCGTAGTGTGCTGGCTCGCTTTGGTGATGATGGAGA
16	TTACAGATTTTCTCCATCATCACCAAAGCGAG
PMR-3’	AATAATGGATCCTTACAGATTTTCTCCATCATCACC
PAC-5’	AATAATGGTACCTTACAGATTTTCTCCATCATCACC

Cloning. Amplification was performed as described in Objective 1. Screening genes and PMR vector were digested with AflIII and BamHI (NEB), and expression genes and PAC vector were digested with AflIII and BanI. Overnight ligation at 16 ºC was followed by background digestion with EcoRI and NdeI for PMR vector and EcoRI and NcoI for PAC vector. The purified plasmids were transformed into electrocompetent DH10B, and recovered cells were plated and incubated overnight. Picked colonies were grown in rich media, and the extracted DNA (Qiagen Miniprep) was sequenced (Genewiz).

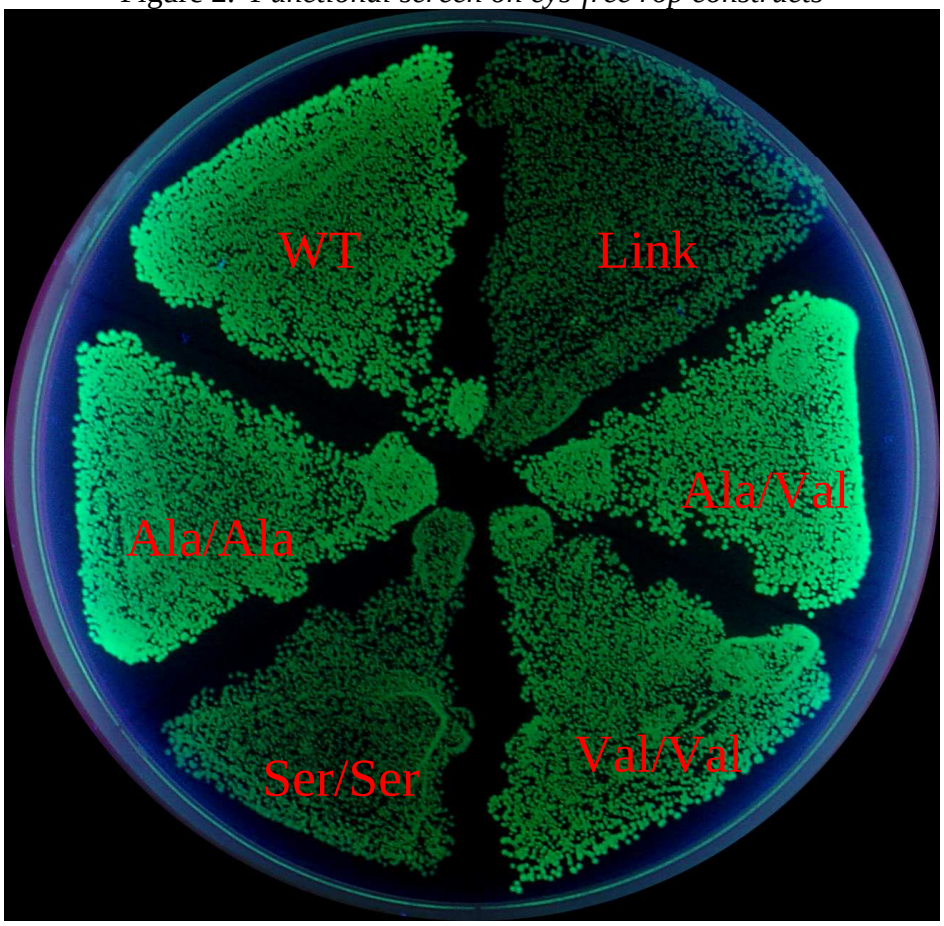
Screening and protein purification were performed as described in Objective 1. CD spectroscopy was performed at 25 μM single-chain monomer.

Results and Discussion

Objective 1: Synthesis and characterization of a cysteine-free rop

Figure 2 shows the results of the functional GFP screen performed on the sythesized cys-free rop constructs. It should be noted that the positive screen was used for this experiment, which displays an inverse phenotype of the negative screen. It is readily seen that only Ser/Ser is not active *in vivo*. “Link” is a negative control which does not contain a rop gene.

Figure 2. *Functional screen on cys-free rop constructs*



Of the active constructs, only Ala/Val (AV) rop was soluble as an overexpressed His_6-fusion protein. CD wavelength scans (Figure 3) show that AV rop has approximately the same amount of α-helical character as wild-type. Additionally, thermal melts (Figure 4) show that AV rop is slightly more stable thermodynamically than wild-type. Interestingly, AV rop is not nearly as stable by chemical denaturation methods (Figure 5). Unfolding in the presence of high urea concentrations is approximately 1 h, while the same conditions take several days for wild-type.

Figure 3. *Wavelength scans of AV and wild-type rop*

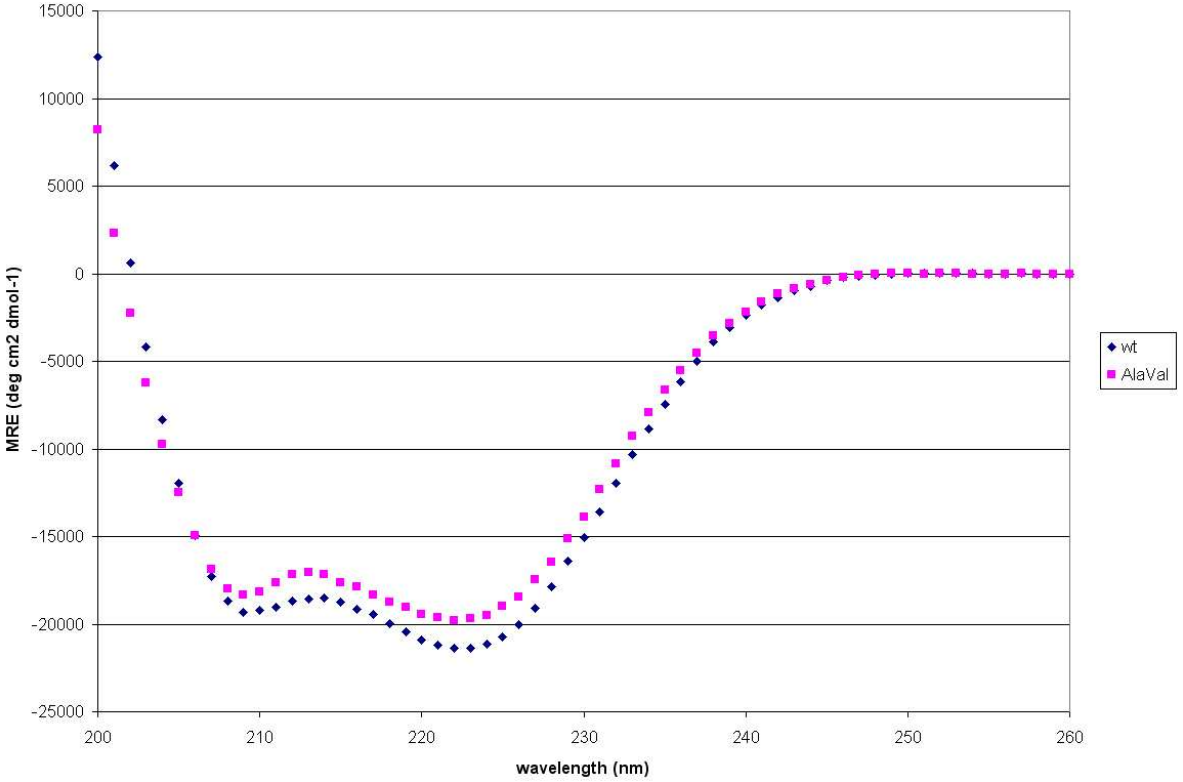


Figure 4. *Thermal melts of AV and wild-type rop*

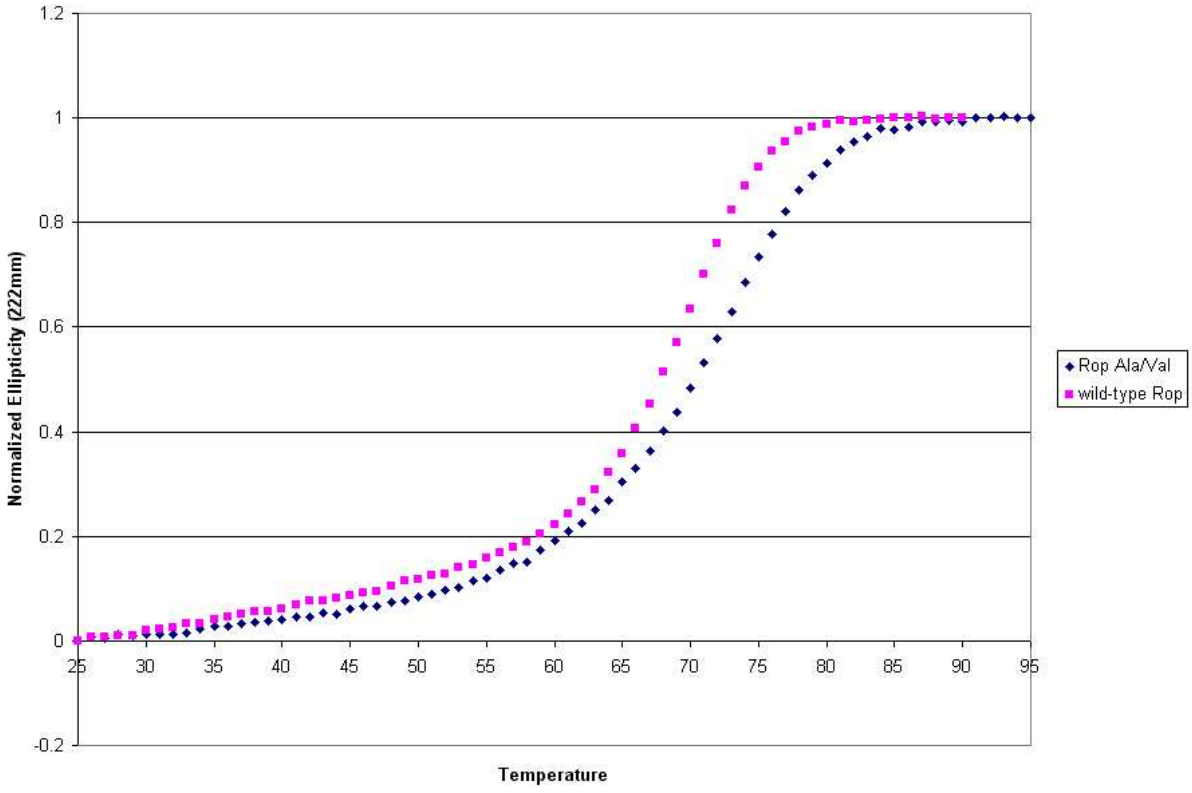
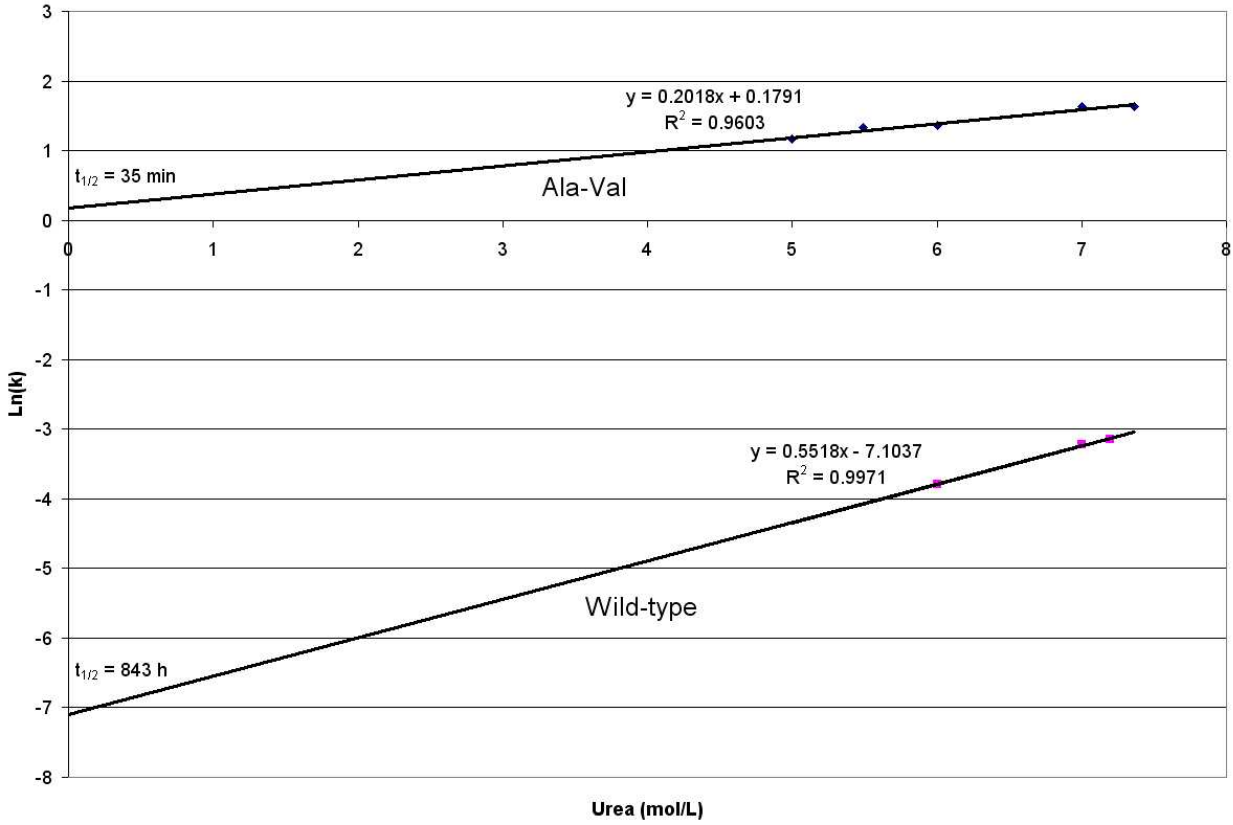
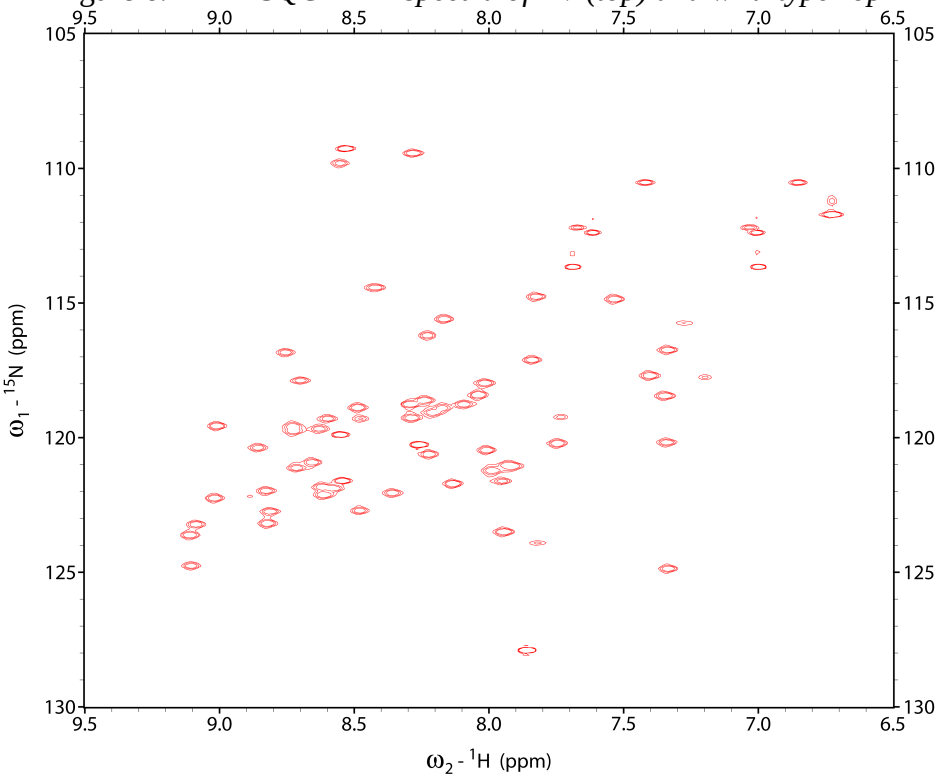


Figure 5. *Rates of unfolding for AV and wild-type rop in urea*



The HSQC NMR spectrum of AV rop (Figure 6) is very similar to wild-type[53] (ITLA). Peaks are clear and well-dispersed, indicating a folded structure.

Figure 6. *^{15}N-HSQC NMR spectra of AV (top) and wild-type rop*



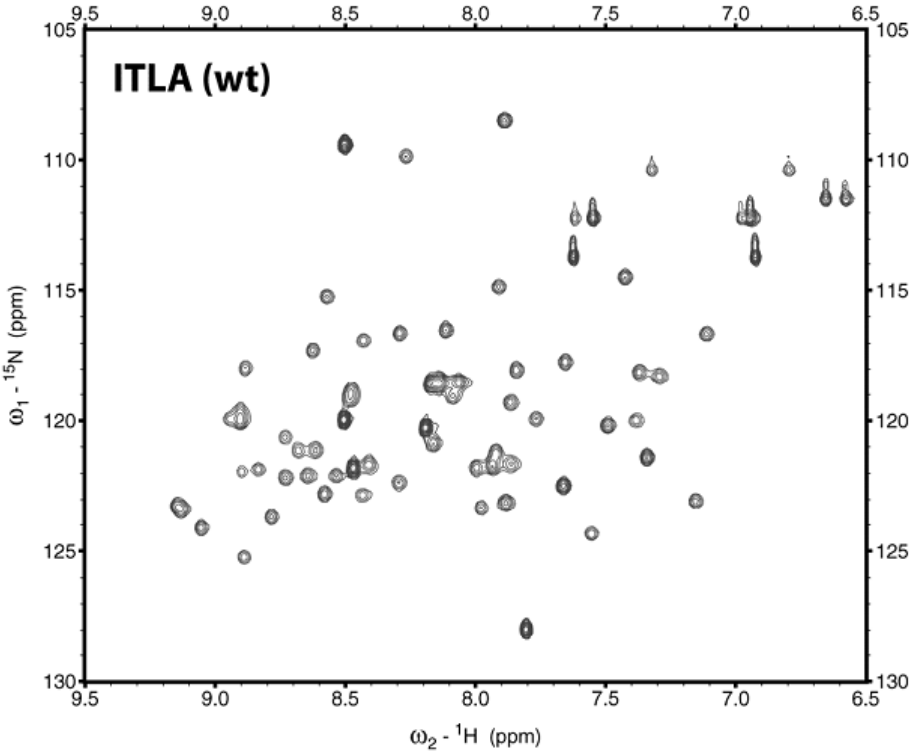


Figure 7 is a sample of the AV crystals obtained with methanol as a precipitant. The prisms were large, and visible deformities did not seem to have any adverse effects on diffraction quality. Figure 8 is the crystal structure of AV rop overlaid on the wild-type structure. The AV structure has not been fully refined at the time of this writing, but the results after molecular replacement phasing and modelling are shown in the figure. The RMSD between the structures is 0.37 Å (as determined by SUPER in Pymol)[54]. Electron density maps show very few areas that are in need of refinement.

Figure 7. *Crystals of AV rop*

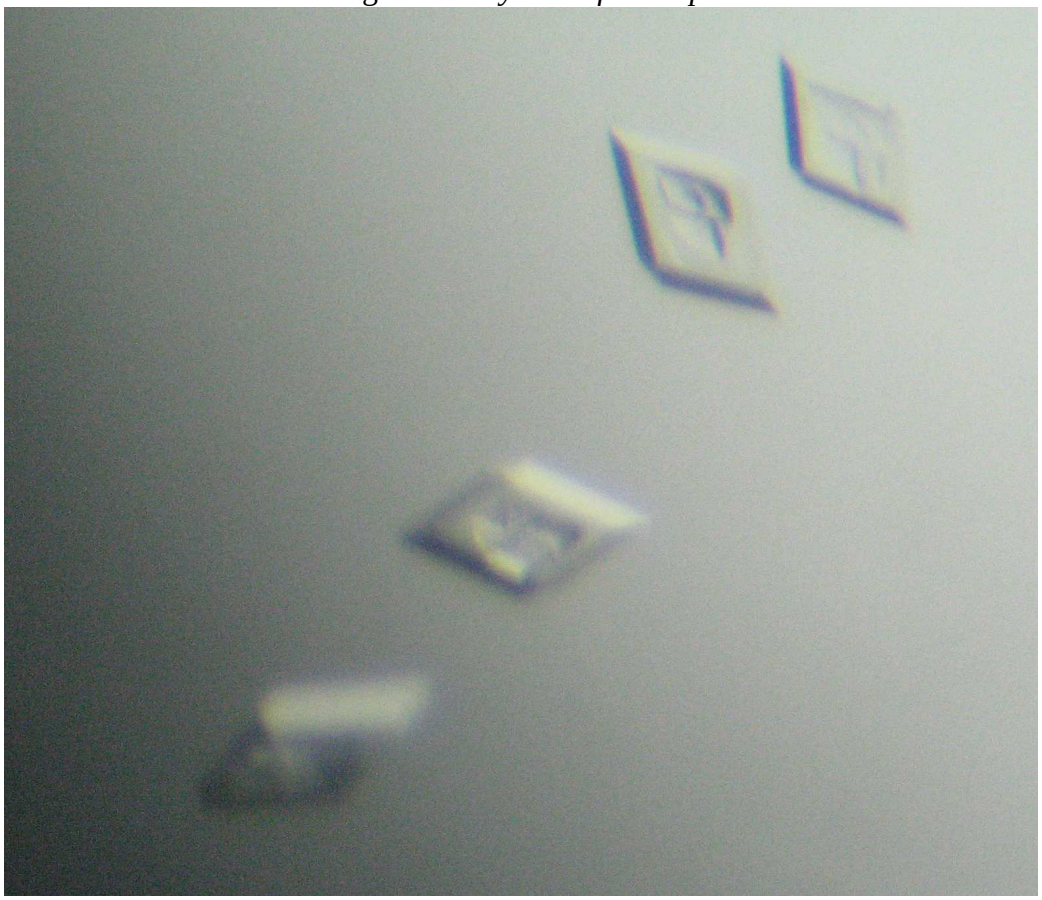
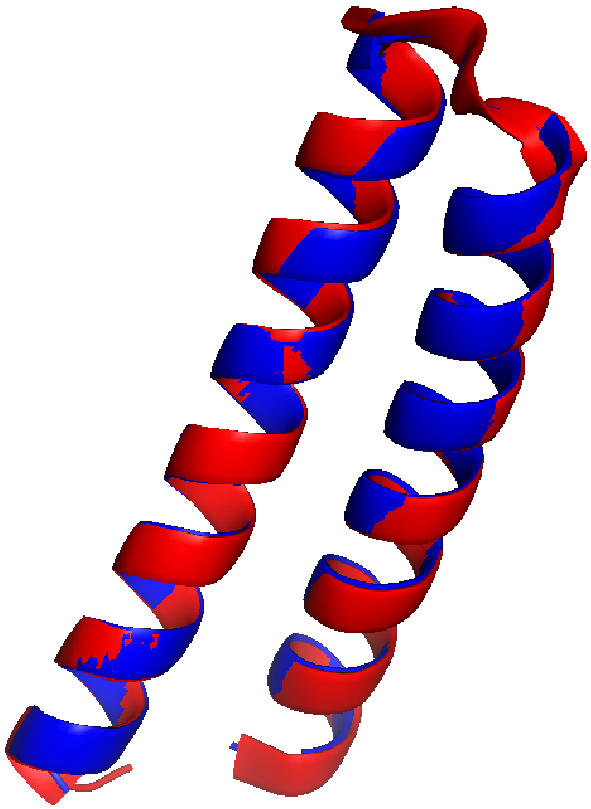


Figure 8. *Crystal structure of AV rop (blue) overlaid on wild-type (red)*



Objective 2: Development and application of a novel high-throughput calorimetric assay

It was determined that enough protein could be purified from 2 mL cultures to perform our High-throughput Calorimetry assay. Additionally, the use of DH10B(DE3) allowed protein overexpression and high plasmid replication from the same cell line. This obviated an additional transformation step for DNA sequencing. Also, the assay seems to produce excellent results. Figure 9 shows three variants screened as controls for the method. LMLL, a highly destabilized variant, binds to the dye immediately. On the other hand, wild-type and AV rop exclude the dye from their hydrophobic cores until they unfold.

Figure 9. *Controls used for High-throughput Calorimetry*

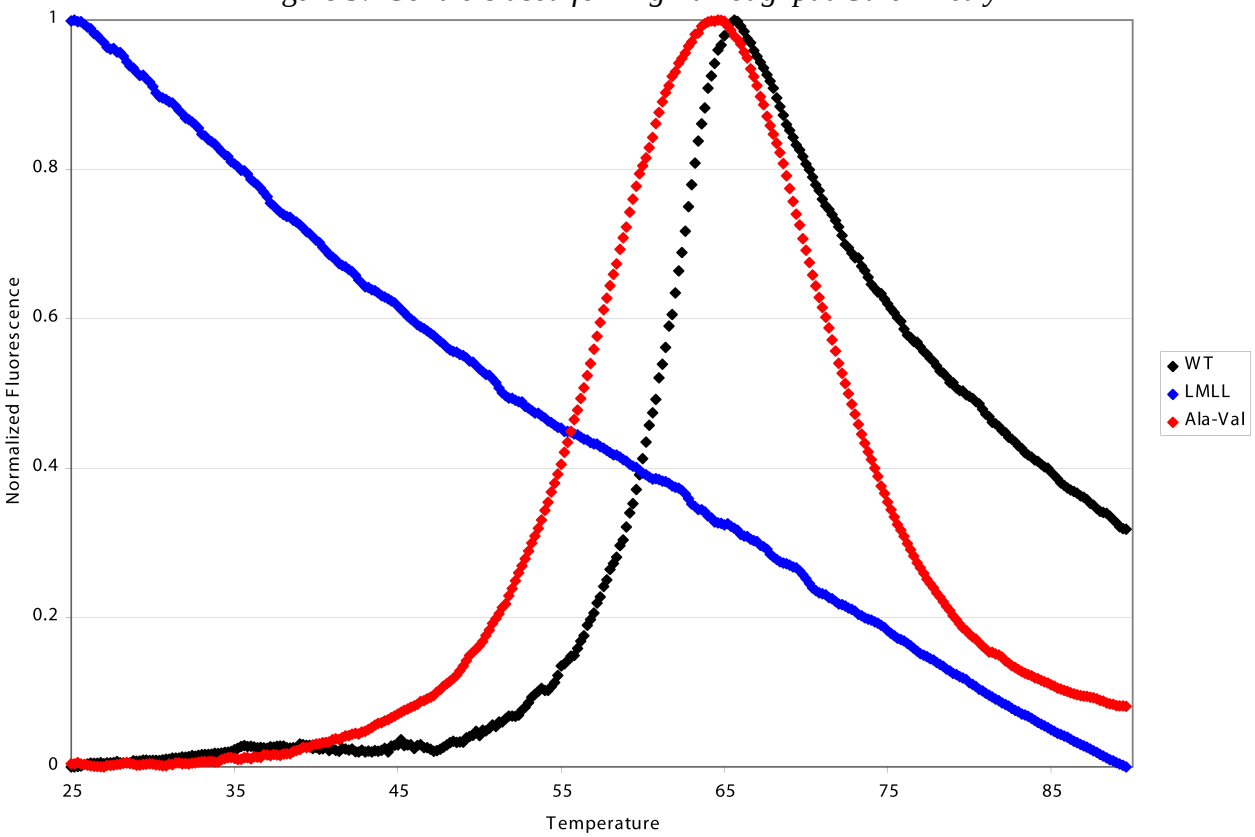
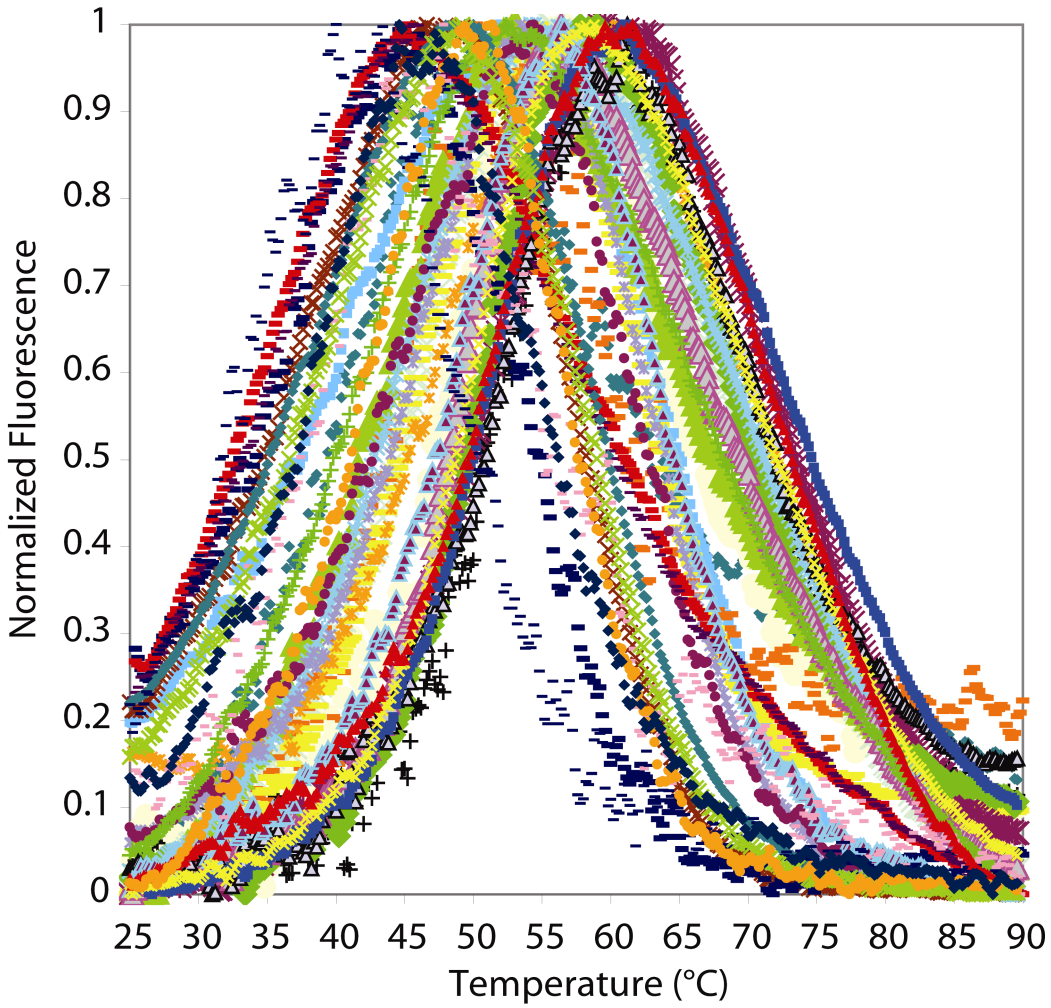


Figure 10 illustrates the throughput of our new calorimetric method; it shows the melt results of dozens of native-like variants from a small rop library which mutates four of the residues in its hydrophobic core.

Figure 10. *High-throughput Calorimetry results from native-like variants in a rop library*



Objective 3: Design and synthesis of single-chain rop variants

Both constructs, MPP and PMP rop, are functionally active, shown by the GFP screen in Figure 11. However, MPP rop appears to be proteolyzed at one of the loop regions, resulting in both full-length and a truncated product (data not shown). Regardless, PMP rop contains a high level of α-helical content, and its urea melt curve indicates cooperativity (Figures 12 and 13). Its thermal melt (Figure 14), however, shows that no refolding occurs, unlike most other native-like dimeric rop constructs.

Figure 11. *In-cell GFP screen of MPP and PMP single-chain rop variants*

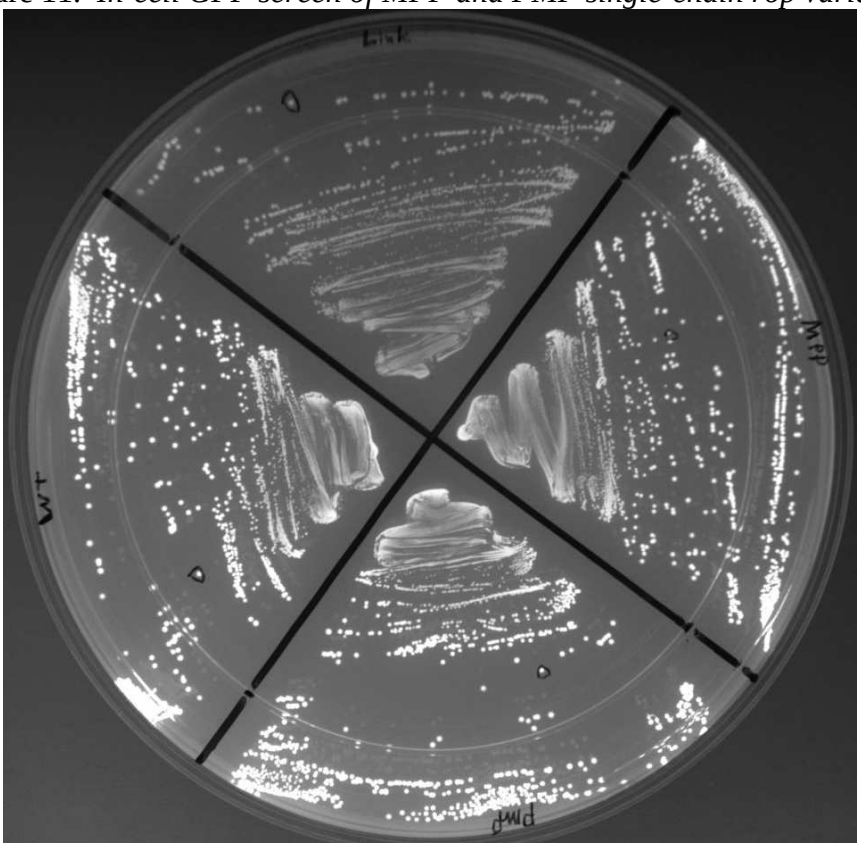


Figure 12. *Wavelength CD scan of PMP rop*

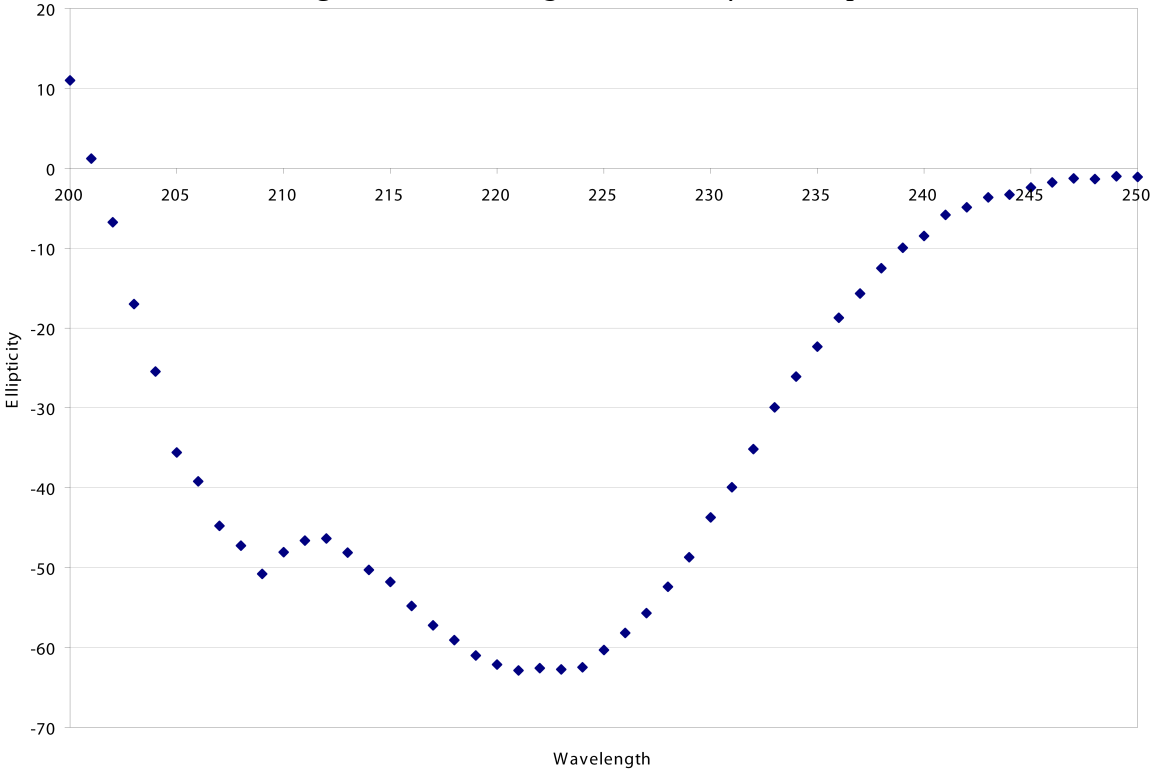


Figure 13. *Urea melt curve of PMP rop*

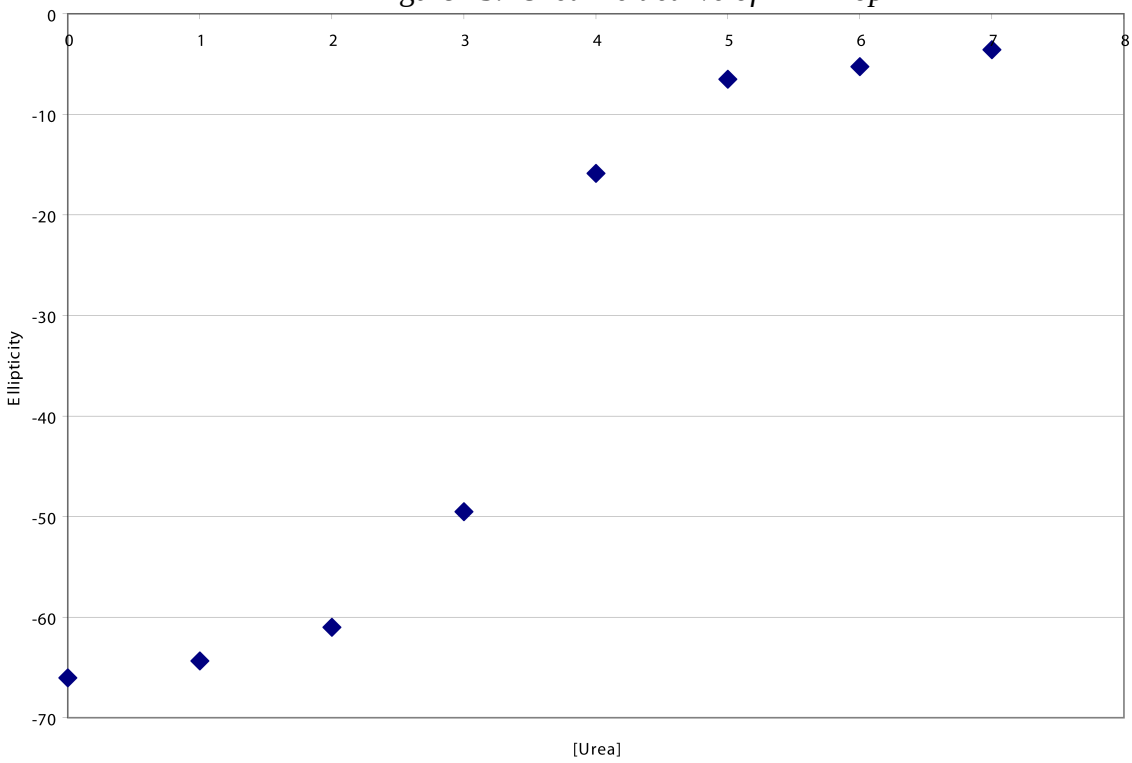
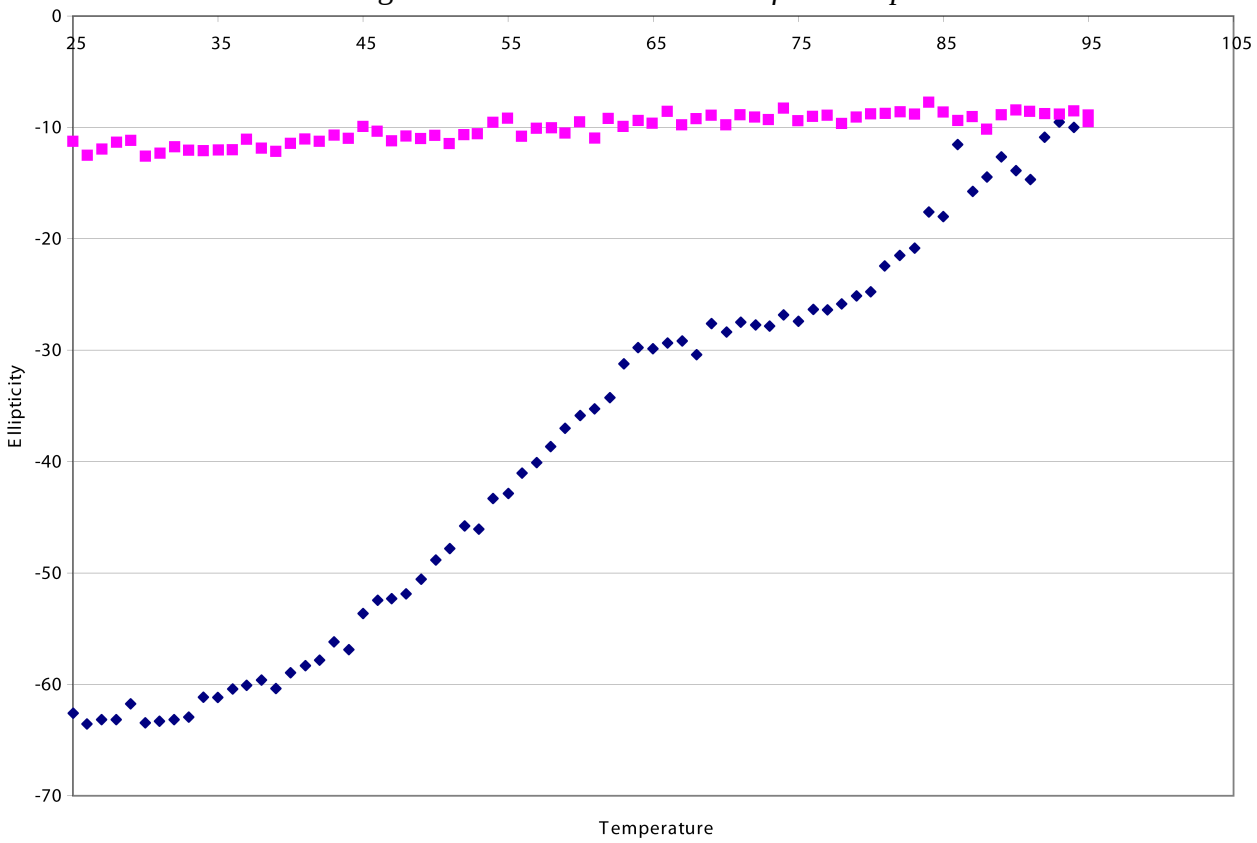


Figure 14. *Thermal melt curve of PMP rop*



Remarks

A novel, cysteine-free variant of wild-type rop has been extensively characterized. High-resolution data shows that this new A38/V52 mutant is very similar to the wild-type structure and maintains its function *in vivo*.

A method for purifying protein variants on a high-throughput scale has been developed and applied to a small library. These variants have also been screened using our newly-developed High-throughput Calorimetry assay, which has shown to be able to characterize proteins rapidly for stability.

Two new single-chain rop variants have been synthesized and characterized to show that both are functionally active *in vivo*, and one has a high degree of α-helical content and shows cooperativity in unfolding.

Future work will likely entail synthesizing single-chain rop libraries which randomize the loop positions in an effort to find more stable variants. These variants will be found by High-throughput Calorimetry and further characterized by NMR spectroscopy and x-ray crystallography.

Bibliography

1. Anfinsen, C. B., The formation and stabilization of protein structure. *Biochem J* **1972,** 128, (4), 737-49.

2. Dill, K. A.; Chan, H. S., From Levinthal to pathways to funnels. *Nat Struct Biol* **1997,** 4, (1), 10-9.

3. Dill, K. A., et al., The protein folding problem: when will it be solved? *Curr Opin Struct Biol* **2007,** 17, (3), 342-6.

4. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990,** 29, (31), 7133-55.

5. Dahiyat, B. I.; Mayo, S. L., Probing the role of packing specificity in protein design. *Proc Natl Acad Sci U S A* **1997,** 94, (19), 10172-7.

6. Baldwin, R. L., Energetics of protein folding. *J Mol Biol* **2007,** 371, (2), 283-301.

7. Jones, D. T., Structural biology. Learning to speak the language of proteins. *Science* **2003,** 302, (5649), 1347-8.

8. Dahiyat, B. I.; Mayo, S. L., De novo protein design: fully automated sequence selection. *Science* **1997,** 278, (5335), 82-7.

9. Harbury, P. B., et al., High-resolution protein design with backbone freedom. *Science* **1998,** 282, (5393), 1462-7.

10. Kuhlman, B., et al., Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003,** 302, (5649), 1364-8.

11. Hecht, M. H., et al., Mutations in lambda repressor's amino-terminal domain: implications for protein stability and DNA binding. *Proc Natl Acad Sci U S A* **1983,** 80, (9), 2676-80.

12. Lim, W. A.; Sauer, R. T., Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* **1989,** 339, (6219), 31-6.

13. Lim, W. A., et al., Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry* **1992,** 31, (17), 4324-33.

14. Bhattacharyya, S., Combinatorial approaches in anticancer drug discovery: recent advances in design and synthesis. *Curr Med Chem* **2001,** 8, (12), 1383-404.

15. Lam, K. S., et al., A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991,** 354, (6348), 82-4.

16. Merrifield, R. B., Solid-phase peptide synthesis. *Adv Enzymol Relat Areas Mol Biol* **1969,** 32, 221-96.

17. Sidhu, S. S., Phage display in pharmaceutical biotechnology. *Curr Opin Biotechnol* **2000,** 11, (6), 610-6.

18. Arndt, K. M., et al., A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble. *J Mol Biol* **2000,** 295, (3), 627-39.

19. Magliery, T. J.; Regan, L., Combinatorial approaches to protein stability and structure. *Eur J Biochem* **2004,** 271, (9), 1595-608.

20. Sondek, J.; Shortle, D., A general strategy for random insertion and substitution mutagenesis: substoichiometric coupling of trinucleotide phosphoramidites. *Proc Natl Acad Sci U S A* **1992,** 89, (8), 3581-5.

21. Moffet, D. A.; Hecht, M. H., De novo proteins from combinatorial libraries. *Chem Rev* **2001,** 101, (10), 3191-203.

22. *Directed evolution library creation : methods and protocols / edited by Frances H. Arnold and George Georgiou*. Humana Press: Totowa, N.J. :, 2003; p x, 224 p. :.

23. Stevens, R. C., High-throughput protein crystallization. *Curr Opin Struct Biol* **2000,** 10, (5), 558-63.

24. Adams, P. D., et al., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **2002,** 58, (Pt 11), 1948-54.

25. Edgell, M. H., et al., High-precision, high-throughput stability determinations facilitated by robotics and a semiautomated titrating fluorometer. *Biochemistry* **2003,** 42, (24), 7587-93.

26. Carver, T. E., et al., Decrypting the biochemical function of an essential gene from Streptococcus pneumoniae using ThermoFluor technology. *J Biol Chem* **2005,** 280, (12), 11704-12.

27. Matulis, D., et al., Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **2005,** 44, (13), 5258-66.

28. Pantoliano, M. W., et al., High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **2001,** 6, (6), 429-40.

29. Lo, M. C., et al., Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* **2004,** 332, (1), 153-9.

30. Ericsson, U. B., et al., Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **2006,** 357, (2), 289-98.

31. Yeh, A. P., et al., Rapid and simple protein-stability screens: application to membrane proteins. *Acta Crystallogr D Biol Crystallogr* **2006,** 62, (Pt 4), 451-7.

32. Vedadi, M., et al., Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* **2006,** 103, (43), 15835-40.

33. Regan, L.; DeGrado, W. F., Characterization of a helical protein designed from first principles. *Science* **1988,** 241, (4868), 976-8.

34. Raleigh, D. P., et al., A De-Novo Designed Protein Mimics the Native-State of Natural Proteins. *J Am Chem Soc* **1995,** 117, (28), 7558-7559.

35. Hill, R. B.; DeGrado, W. F., Solutions structure of alpha D-2, a nativelike de novo designed protein. *J Am Chem Soc* **1998,** 120, (6), 1138-1145.

36. Cesareni, G., et al., Control of ColE1 DNA replication: the rop gene product negatively affects transcription from the replication primer promoter. *Proc Natl Acad Sci U S A* **1982,** 79, (20), 6313-7.

37. Banner, D. W., et al., Crystallization of the ColE1 Rop protein. *J Mol Biol* **1983,** 170, (4), 1059-60.

38. Banner, D. W., et al., Structure of the ColE1 Rop protein at 1.7 .ANG. resolution. *J Mol Biol.* **1987,** 196, (3), 657-75.

39. Eberle, W., et al., The structure of Co1E1 rop in solution. *J Biomol NMR* **1991,** 1, (1), 71-82.

40. Castagnoli, L., et al., Genetic and structural analysis of the ColE1 Rop (Rom) protein. *Embo J* **1989,** 8, (2), 621-9.

41. Munson, M., et al., Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci* **1994,** 3, (11), 2015-22.

42. Predki, P. F.; Regan, L., Redesigning the topology of a four-helix-bundle protein: monomeric Rop. *Biochemistry* **1995,** 34, (31), 9834-9.

43. Kresse, H. P., et al., Four-helix bundle topology re-engineered: monomeric Rop protein variants with different loop arrangements. *Protein Eng* **2001,** 14, (11), 897-901.

44. Magliery, T. J.; Regan, L., A cell-based screen for function of the four-helix bundle protein Rop: a new tool for combinatorial experiments in biophysics. *Protein Eng Des Sel* **2004,** 17, (1), 77-83.

45. Stemmer, W. P., et al., Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **1995,** 164, (1), 49-53.

46. Kapust, R. B.; Waugh, D. S., Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr Purif* **2000,** 19, (2), 312-8.

47. Pflugrath, J. W., The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* **1999,** 55, (Pt 10), 1718-25.

48. McCoy, A. J., Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **2007,** 63, (Pt 1), 32-41.

49. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** 60, (Pt 12 Pt 1), 2126-32.

50. Vagin, A. A., et al., REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **2004,** 60, (Pt 12 Pt 1), 2184-95.

51. Chae, Y. K., et al., Protein production by stationary phase induction (SPI). *Protein Pept Lett* **2003,** 10, (4), 369-74.

52. Gao, X., et al., Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res* **2003,** 31, (22), e143.

53. Magliery, T. J.; Regan, L., An experimental approach to the ‘inverse' folding problem: combinatorial exploration of the core of a four-helix bundle protein. (manuscript in preparation).

54. DeLano, W. L. *The PyMOL Molecular Graphics System*, DeLano Scientific: Palo Alto, CA, USA, 2002.